

USE OF ONLINE DATABANKS

FOR INFORMATION RETRIEVAL

---

N. BASKEVITCH

LTR INDUSTRIES

Presented to the joint session of Technology/Smoke study groups.

- Florence 1983 -

2001206477

USE OF ON LINE  
DATA BANK FOR INFORMATION RETRIEVAL

---

INTRODUCTION

The basis of all scientific work, in the tobacco or in any other field, is a complete revue of what has been done before. It is then very important to have access to good tools for literature survey.

We all know how this is difficult, and specially in the tobacco field where informations are scattered in a number of magazines. What possibilities can use research people :

- Or their company or institute has a library. It must be of a very large size to catch all information, and this is expensive, in surface as in cost of magazines,
- Or they are lucky to work near from a university which has a large library,
- Or they have access to data bank terminals through which they can be in contact with most of the information stored in libraries in the entire world, and this is the subject I would like to talk to you about today.

.../...

2001206478

Among the numerous tools that computers are giving to us nowadays, there is one of direct interest for scientists : on-line databanks for information retrieval.

A databank is a computer system which gives access to information. Generally speaking, there are two types of databanks :

- 1 - Documentation databanks - They contain literature references about various subjects, from Chemistry (Chemical Abstracts, for example) to economy (Economical Abstracts) or patents (INPI, PATSEARCH).

Some databanks can be specialized in a specific subject like paper (PIRA) or coffee (COFFEELINE).

In theory, these databanks guarantee a complete survey of the information available in a given field.

The user can be confident that specialists have stored in the databank all papers published in a given list of magazines.

- 2 - Facts databanks - This second type of databanks gives facts, for example thermodynamical values, meteorological informations, commercial statistics or information on financial relationships between companies.

.../...

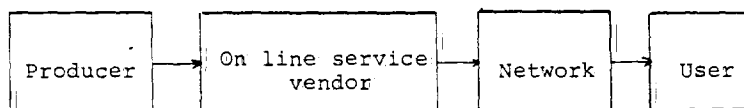
2001206479

Those databanks answer an always increasing demand for information retrieval. They allow the replacement of large surface libraries storing collections of periodicals.

By the end of 1982, it was possible to have access in France, to 340 different databanks. By the middle of 1983, a new list of 45 databanks has been added.

How to communicate with a databank ?

The process is simple when you look at it as a whole. When you get to details it is a little more complicated.



The PRODUCER collects, screens, updates and analyse informations linked with the subject. He runs a computer treatment and prepare a computer tape which he forwards at regular intervals to the ONLINE SERVICE VENDOR.

The ONLINE SERVICE VENDOR owns a powerful computer. He loads and stores on this computer the databanks obtained from the PRODUCER. He assures the good run of the questionning computer and of the software.

.../...

2001206480

Some example of PRODUCERS and ONLINE SERVICE VENDORS :

CHEMICAL ABSTRACTS is the PRODUCER of a databank called CA SEARCH.

This databank is loaded on computers belonging to different ONLINE SERVICE VENDORS like :

- . DIALOG , of Lockheed Missiles and Space Company in Palo Alto ;
- . SDC , of System Development Corporation, USA ;
- . INFOLINE, of Pergamon in the U.K. ;
- . ESA-IRS , of the European Space Agency in Italy ;
- . QUESTEL , of Telesystem in France.

Between the computer terminal owned by the USER and the ONLINE SERVICE VENDOR computer, one needs a wire which put into contact two telephone stations, wherever they are in the world.

However the cost of long distance calls rises strongly the price of the communication. TELECOM services in many countries have decided to set up specialised networks to carry over information between computers and terminals. The use of these networks is considerably cheaper than "normal" long distance calls.

.../...

2001206481

The best known NETWORKS are :

- in France : TRANSPAC
- in Europe : EURONET
- in the USA : TYMNET or TELENET
- in South Africa : SAPONET
- in Brazil : INTERDATA

New networks are created at a high rate. This will make easier the access of users to databanks and contribute to their development.

Finally, the USER, to get in touch with the computer of the ONLINE SERVICE VENDOR needs a terminal. This can be either a keyboard-screen, a keyboard-printer or a keyboard-screen-printer combination. This terminal is equipped with an acoustic or electric modem to be connected with the telephone line.

How does the system work ?

- The USER has a terminal and a telephone line,
- He asks the Network to supply him with an identification number (NUI).
- He signs a contract with the ONLINE SERVICE VENDOR who gives him a confidential access number to his computer.
- To question a databank, the USER operates as follows :
  - . Dial the phone number of the NETWORK, then the terminal is connected and all the dialog is now made from the keyboard.

.../...

2001206482

. Type then in order :

- the phone number of the ONLINE SERVICE VENDOR,
- the identification number (NUI),
- the confidential access number to the ONLINE SERVICE VENDOR,
- the code name of the desired databank.

Everything is now ready for questionning the databank :

Let us take an example :

We would like to run a litterature survey on the analysis of carbon monoxide in cigarette smoke.

We have decided to question CHEMICAL ABSTRACTS loaded on ESA-IRS computer.

We have ddaled the Network TRANSPAC Number, the ESA-IRS phone number, one NUI and one confidential password. The computer accepts the connection and gives the starting time.

We ask to question CHEMICAL ABSTRACTS Databank, and will use keywords to obtain one information.

For each keyword the computer will answer the number of appearance in the databank. For each of the following keywords the answer is :

- carbone	answer : 250362 appearance
- monoxide	28709 appearance
- smoke	6955 appearance
- cigarette	494 appearance
- analyse	320746 appearance

.../...

2001206483

Then we combine the five keywords together looking for papers having all the keywords at the same time. Computer answers six papers.

We ask to look at the titles and reference of the six articles (see Annex 1).

Interrogation is then ended and the communication disconnected.

#### Let us talk about cost

The price of a terminal with screen, printer and modem is in the range of 15 000 to 30 000 FF. This can be rented for around 1 000 FF per month.

The use of the NETWORK costs from 30 FF to 40 FF an hour.  
The use of a databank, invoiced by the ONLINE SERVICE VENDOR costs between 300 and 400 FF per hour.

The incoming on the market of small terminals of the type MINITEL, rented by the telephone services for around 70 FF per month will considerably increase in the near future the number of users of databanks.

#### International Networks

The number of countries having access to each other through NETWORKS specifically for computer communication is increasing every month and will soon cover the largest part of the world.

The annexed list (Annex 2) gives the today situation of Networks connecting France with other countries.

Among the members of CORESTA, 117 on 136 will belong to countries interconnected with France and Europe by the end of 1984.

.../...

2001206484



### LANGUAGES

Each ONLINE SERVICE VENDOR has developped its own language to load the databanks on its computer. The USER needs to have a certain natural talent to cope with the different languages.

If you want to question, from a European country, most of the databanks available around the world, you need, besides English, to know at least 4 languages :

- DIALOG (Lockheed)
- ORBIT (SDC)
- RECON (ESA-IRS)
- MISTRAL (QUESTEL)

These languages have many commun features, but, and this is more cumbersome, a number of small differences.

ONLINE SERVICE VENDORS propose training sessions and handbooks to learn the languages. However we must say that there is a need for some practise before a language can be mastered, in order not to loose time during an interrogation.

### HOW TO SELECT THE DATA BANK

Databanks are a relatively new phenomenon, but many databanks are already available. There is a business competition between different ONLINE SERVICE VENDORS. Each of them tries to offer the most complete and complementary range of databanks so as to gain regular customers who will appreciate to avoid the use of different languages at one time.

.../...

2001206485

The beginner user will start by questioning different ONLINE SERVICE VENDORS offering similar possibilities. Sooner or later he will find that some databanks are richer in information than others and rather than to question different databanks on one given subject, he will systematically call on the richest.

As far as tobacco is concerned, the databank produced by CHEMICAL ABSTRACTS, available through several ONLINE SERVICE VENDORS, is, second on our mind, the richest currently available, at least for technology and chemistry aspects. But Agronomy is not as well covered.

To illustrate this question, I decided to run the following exercise : I took literature references in CORESTA bulletins and in Tobacco Abstracts ; and we tried to retrieve them in different databanks.

The title of the first page was :

"Chemical studies on tobacco smoke"

LI : Studies on non volatile Nitrosamine in tobacco

Authors : S.S. HECHT .... and T.C. TSO

Reference : Beiträge zur Tabakforschung 9 1-7 (1977)

Different databanks were questioned, using the following keywords :

- Tobacco
- Smoke
- Nitrosamine
- Non volatile

.../...

2001206486

We only found this paper in CHEMICAL ABSTRACTS databank, using the ONLINE SERVICE VENDOR "ESA-IRS".

In this bank, 4 references corresponded to the combination of the 4 keywords, and among them, the paper which we were looking for.

We operated in the same way for a second paper .

Title : Biochemical Changes in leaf pigments and chemical constituents during Flue-curing of tobacco.

Authors : A. GOPALAM and N.C. GOPALACHARI

Reference : Tobacco Research 5 117-124 (1979)

We combined 4 keywords :

- Biochemical
- Changes
- Leaf
- Pigment

Again, we only found this paper in CHEMICAL ABSTRACTS databank where 4 papers corresponded to the 4 keywords.

But we were unable to find trace of several papers on tobacco, published in a Canadian magazine "THE LIGHTER" or in Japanese magazines.

.../...

2001206487

The first objective of this presentation was to describe for those of the audience who were not already familiar with it, a new information retrieval technique through the use of online databanks.

- Do you think that the creations of a databank, which could be called CORESTA databank could be useful to you ?

A CORESTA databank could also supply facts, for example the survey of BLUE MOLD progress.

What I suggest today is, as a first step, to set up a task force, in the Technology group which will run a feasibility study on the development of a CORESTA databank.

We will obviously consult other associations, comparable to CORESTA, which have already decided to go in this direction.

or

RESEARCH ORGANIZATION FOR THE PAPER AND BOARD (PIRA) in the U.K.